Seres humanos controlando la IA - Expectativas regulatorias emergentes en el sector financiero

Perspectivas de FSI | No. 35 | 03 agosto 2021

por  [Jeffery Yong](https://www.bis.org/author/jeffery_yong.htm) y [Jeremy Prenio](https://www.bis.org/author/jermy_prenio.htm)

[**PDF texto completo** (383kb)](https://www.bis.org/fsi/publ/insights35.pdf)  | 28 páginas

* [**Resumen ejecutivo**](https://www.bis.org/fsi/publ/insights35_summary.pdf)**(90 KB, PDF)**

La inteligencia artificial (IA) tiene el potencial de mejorar significativamente la prestación de servicios financieros. Varias autoridades financieras han comenzado recientemente a desarrollar marcos, describiendo sus expectativas sobre la gobernanza y el uso de la IA por parte de las instituciones financieras. Estos marcos convergen en principios rectores comunes sobre confiabilidad, rendición de cuentas, transparencia, equidad y ética. En general, los requisitos de modelado, gestión de riesgos y gobernanza de alto nivel existentes para los modelos tradicionales ya cubren estos principios de IA. La diferencia clave entre los requisitos reglamentarios para los modelos tradicionales y los de IA es el mayor énfasis de estos últimos en las responsabilidades humanas para evitar la discriminación y otras decisiones no éticas. Además, si bien los principios emergentes de la IA son útiles, cada vez hay más llamados a los reguladores financieros para que brinden una guía práctica más concreta dados los desafíos en la implementación de los principios. Estos desafíos incluyen la velocidad y la escala de la adopción de IA por parte de las instituciones financieras, mayores puntos de contacto con problemas éticos y de equidad, la construcción técnica de los algoritmos de IA y la falta de explicabilidad del modelo. Estos desafíos también exigen una respuesta regulatoria y de supervisión proporcional y coordinada. A medida que surjan enfoques regulatorios y prácticas de supervisión más específicos, los organismos globales de establecimiento de estándares podrían estar en una mejor posición para desarrollar estándares en esta área. construcción técnica de algoritmos de IA y falta de explicabilidad del modelo. Estos desafíos también exigen una respuesta regulatoria y de supervisión proporcional y coordinada. A medida que surjan enfoques regulatorios y prácticas de supervisión más específicos, los organismos globales de establecimiento de estándares podrían estar en una mejor posición para desarrollar estándares en esta área. construcción técnica de algoritmos de IA y falta de explicabilidad del modelo. Estos desafíos también exigen una respuesta regulatoria y de supervisión proporcional y coordinada. A medida que surjan enfoques regulatorios y prácticas de supervisión más específicos, los organismos globales de establecimiento de estándares podrían estar en una mejor posición para desarrollar estándares en esta área.

Clasificación JEL: C60, G29, G38, O30.

Palabras clave: inteligencia artificial, aprendizaje automático, gobierno corporativo, gestión de riesgos, modelado de riesgos.





Humanos que mantienen la IA bajo control: expectativas regulatorias emergentes en el sector financiero1

Resumen Ejecutivo

La inteligencia artificial (IA), incluido el aprendizaje automático (ML), puede ofrecer un potencial significativo para mejorar la prestación de servicios financieros y los procesos operativos y de gestión de riesgos. La tecnología es ahora parte integrante de los servicios financieros y no hay duda de que continuará impulsando cambios profundos para los consumidores y las instituciones financieras. Esta tendencia ha sido apoyada por los esfuerzos de las autoridades financieras para promover la innovación y el uso de nuevas tecnologías en el sector financiero. Al hacerlo, los marcos regulatorios sólidos son esenciales para optimizar los beneficios y minimizar los riesgos de estas nuevas tecnologías.

Existen marcos o principios de gobernanza de la IA que se aplican en todas las industrias y, más recientemente, varias autoridades financieras han iniciado el desarrollo de marcos similares para el sector financiero. Dentro de estos marcos, varios temas comunes convergen en principios rectores generales sobre confiabilidad, rendición de cuentas, transparencia, equidad y ética. Otros principios rectores mencionados en ciertos marcos se relacionan con la privacidad de los datos, la dependencia de terceros y la resiliencia operativa. Si bien estos principios de alto nivel son útiles para proporcionar una indicación amplia de lo que las empresas deben considerar al utilizar tecnologías de IA, hay crecientes llamamientos para que los reguladores financieros proporcionen una orientación práctica más concreta. Un enfoque para satisfacer esta necesidad de la industria es que los reguladores proporcionen una compilación de las mejores prácticas emergentes de la industria sobre la gobernanza de la IA, cuando estén disponibles, para cada uno de estos principios generalmente aceptados.

Los requisitos existentes sobre gobernanza, gestión de riesgos, así como desarrollo y operación de modelos tradicionales también se aplican a los modelos de IA. Estos incluyen requisitos de gobernanza que responsabilizan el uso de tales modelos en los consejos de administración y la alta dirección de las instituciones financieras. Más específicamente, las empresas generalmente deben contar con procesos sólidos de validación de modelos para determinar la confiabilidad de los resultados de la modelización. Es importante destacar que los supervisores esperan que dichos modelos sean transparentes, no solo como parte de las buenas prácticas de gestión de riesgos, sino también para permitir la revisión supervisora de los modelos. Además, las leyes, normas u orientaciones reglamentarias existentes abarcan la privacidad de los datos, la dependencia de terceros y la resiliencia operativa, incluso en el uso de modelos.

Si bien la mayoría de los problemas que surgen del uso de la IA por parte de las instituciones financieras son similares a los de los modelos tradicionales, la perspectiva podría ser diferente. En el contexto de los modelos de IA, algunos de los principios rectores comunes identificados anteriormente se consideran desde la perspectiva de la equidad. Por ejemplo, garantizar la fiabilidad/ solidez de los modelos de IA tiene como objetivo evitar causar discriminación debido a decisiones inexactas. Además, garantizar la rendición de cuentas y la transparencia en el uso de la IA incluye asegurarse de que los interesados conozcan las decisiones basadas en datos y tengan canales para indagar y desafiar estas decisiones.

El mayor énfasis en la equidad en el uso de la IA da como resultados llamados a una mayor intervención humana. El impulso general de los pronunciamientos de IA por parte de los organismos reguladores parece centrarse en resultados indeseables, incluido el sesgo no intencionado que conduce a resultados discriminatorios que pueden surgir de las automatizaciones y la falta de transparencia en los modelos de IA. Aunque el modelado basado en humanos puede sufrir debilidades similares de los modelos de IA, por ejemplo, sesgos o errores, una característica distintiva clave del primero es que los humanos pueden ser considerados inequívocamente responsables. El uso de la IA, sin embargo, puede conducir a la reflexión filosófica de la demarcación entre las máquinas y los seres humanos. Este desafío se ve agravado por una serie de factores, entre ellos i) la velocidad y la escala de la adopción de la IA por parte de las instituciones financieras; ii) construcción técnica de algoritmos de IA; y iii) falta de explicabilidad del modelo. Desde un punto de vista regulatorio, responsabilizar firme y claramente a las personas responsables dentro de una empresa es clave para poner en práctica marcos regulatorios sólidos de IA. Sin embargo, será necesario hacer concesiones entre cosechar los beneficios de la automatización de máquinas a gran escala contra la necesidad de aportes humanos y supervisión.

Hay margen para definir aún más la equidad para apoyar una gobernanza sólida de la IA. La equidad, como se describe en los documentos de política cubiertos en el documento, se relaciona con evitar resultados discriminatorios. Sin embargo, la no discriminación puede no ser explícita en las leyes de protección al consumidor en algunas jurisdicciones. Hacer explícitos los objetivos de no discriminación puede ayudar a proporcionar una buena base para definir la equidad en el contexto de la IA, proporcionar una base legal para que las autoridades financieras emitan orientaciones relacionadas con la IA y, al mismo tiempo, garantizar que las decisiones impulsadas por la IA, basadas en modelos tradicionales y humanas en los servicios financieros se evalúen con arreglo al mismo estándar.

Los desafíos y la complejidad que presenta la IA exigen una respuesta regulatoria y supervisora proporcional y coordinada. Esto requiere diferenciar el tratamiento regulatorio y supervisor sobre el uso de modelos de IA, dependiendo de la conducta y los riesgos prudenciales que plantean. Los modelos de IA cuyos resultados tengan implicaciones significativas en la conducta y los riesgos prudenciales deberán estar sujetos a un tratamiento regulatorio y supervisor más estricto que aquellos con implicaciones menos significativas. Además, el uso de la IA por parte de las instituciones financieras tendrá implicaciones para la rentabilidad, el impacto en el mercado, la protección del consumidor y la reputación. Esto requiere una mayor coordinación entre las autoridades prudenciales y de conducta en la supervisión del despliegue de la IA en los servicios financieros.

Dados los nuevos temas comunes sobre la gobernanza de la IA en el sector financiero, parece haber margen para que los organismos de normalización financiera elaboren orientaciones o normas internacionales en este ámbito. Las opiniones de las autoridades sobre la forma en que deben aplicarse estos temas comunes siguen evolucionando. Un intercambio continuo de opiniones y experiencias a nivel internacional podría conducir eventualmente a la elaboración de normas internacionales. Tales estándares internacionales podrían ser útiles, particularmente para las jurisdicciones que recién están comenzando su viaje de transformación digital. También pueden servir como un punto de referencia mínimo para guiar el despliegue ordenado de las tecnologías de IA dentro del sector financiero. A medida que surjan enfoques regulatorios más específicos o expectativas de supervisión sobre aspectos específicos de los casos de uso de IA, los organismos normativos pueden identificar esas "mejores prácticas" comunes que serán útiles para que otras jurisdicciones las consideren. Al mismo tiempo, dada la evolución de las tendencias tecnológicas, la orientación basada en principios sigue teniendo sus beneficios y puede complementar ese enfoque basado en las mejores prácticas.

Sección 1 – Introducción

1. La inteligencia artificial (IA), incluido el aprendizaje automático (ML), es una de las tecnologías definitorias que está remodelando el sector financiero. Dentro de los círculos regulatorios, una definición comúnmente citada de IA es la del Consejo de Estabilidad Financiera (2017), que establece que "la aplicación de herramientas computacionales para abordar tareas que tradicionalmente requieren sofisticación humana se denomina en términos generales 'inteligencia artificial'". El documento describió además que el ML se puede definir como un "método para diseñar una secuencia de acciones para resolver un problema, conocido como algoritmos, que se optimizan automáticamente a través de la experiencia y con una intervención humana limitada o nula". Al igual que la forma en que Internet transformó la forma en que realizamos operaciones bancarias o compramos seguros, la IA tiene el potencial de mejorar significativamente la prestación de servicios financieros, pero también trae nuevos riesgos con los que los supervisores del sector financiero deben lidiar.2

2.AI tecnología puede mejorar significativamente la prestación de servicios financieros a los consumidores, así como los procesos operativos y de gestión de riesgos dentro de las empresas.3 Ejemplos de casos de uso de IA en términos de beneficios para el consumidor incluyen:

• ampliar el acceso al crédito

• robo-asesores que proporcionan asesoramiento de inversión automatizado basado en los objetivos de inversión y los perfiles de riesgo de un consumidor

• chatbots que proporcionan respuesta instantánea a las consultas básicas de los clientes

• Procesamiento de reclamaciones de seguros más eficiente Internamente dentro de las empresas, la tecnología de IA promete un enorme potencial en las siguientes áreas:

• identificación de transacciones financieras sospechosas que podrían ser fraudulentas, lavado de dinero o financiamiento del terrorismo

• puntuación de riesgo que permite la concesión automatizada de préstamos (en algunos casos incluso sin garantía) o la decisión de precios de seguros

• gestión de riesgos y/o cálculo de los requisitos de capital reglamentario

3. Aunque la IA puede beneficiar tanto a los consumidores como a las empresas, también puede introducir y/o exacerbar las exposiciones al riesgo. Los riesgos, como el sesgo no intencionado o la discriminación contra determinados grupos de consumidores, no solo son una cuestión de conducta de mercado o de protección de los consumidores, sino que también afectan a los supervisores prudenciales cuando dichos riesgos se traducen en exposiciones financieras para las empresas o si dan lugar a riesgos operativos a gran escala, incluidos el riesgo cibernético y el riesgo reputacional. Además, los riesgos prudenciales pueden surgir de la subvaloración a gran escala de los productos financieros o de errores sistemáticos en la suscripción de nuevos consumidores financieros. En última instancia, deben existir salvaguardias para proteger los intereses de los consumidores y mantener la seguridad y la solidez de las instituciones financieras.

4. A medida que más instituciones financieras están aumentando el uso de la IA para respaldar sus procesos comerciales, los reguladores financieros están comenzando a establecer o actualizar marcos regulatorios específicos sobre la gobernanza de la IA. Con algunas excepciones, como en la Unión Europea (UE), donde ya existe una propuesta legislativa para armonizar las normas para la IA, la mayoría de los marcos aún se encuentran en las primeras etapas de desarrollo y van desde la aplicación de los requisitos de gobernanza empresarial basados en principios existentes en un contexto de IA hasta guías prácticas de supervisión no vinculantes sobre cómo gestionar los riesgos de gobernanza de la IA. En varios casos, estos marcos complementan o se cruzan con marcos intersectoriales sobre gobernanza de la IA desarrollados por reguladores no financieros, como las autoridades de ciberseguridad y/o protección de datos. Si bien algunos pueden argumentar que, desde un punto de vista regulatorio, el uso de la IA no es nada nuevo (por ejemplo, es comparable a la adopción de modelos internos por parte de las empresas), la escala y la velocidad de la adopción de la IA merecen una atención regulatoria especial. Contar con marcos de gobernanza de IA sólidos es cada vez más importante dada la adopción cada vez más amplia por parte de las empresas y el creciente número de autoridades financieras que promueven la innovación y la tecnología en el sector financiero. La innovación financiera no debe comprometer los mandatos básicos de los supervisores del sector financiero.

5. Este documento examina una selección de documentos de política (Tabla 1) sobre gobernanza de la IA emitidos por autoridades financieras o grupos formados por ellas en nueve jurisdicciones y otras directrices de gobernanza de IA intersectoriales que se aplican al sector financiero. El documento tiene como objetivo proporcionar una instantánea de los enfoques regulatorios existentes sobre la gobernanza de la IA, incluido el pensamiento inicial expresado en los documentos de consulta, e identificar temas regulatorios comunes emergentes, incluida la orientación general de IA intersectorial relevante. El ejercicio de evaluación comparativa de políticas se complementa con entrevistas con cuatro autoridades financieras.4 Otras autoridades reguladoras pueden inspirarse en este balance para considerar la adopción de enfoques regulatorios relevantes que sean apropiados para sus circunstancias locales.



6. El resto del documento está estructurado de la siguiente manera. La sección 2 resume las expectativas de determinadas autoridades sobre la gobernanza de la IA, mientras que la sección 3 destaca la pertinencia de las normas internacionales existentes en este contexto. La Sección 4 describe los desafíos en la implementación de marcos de gobernanza sólidos en torno al uso de la IA, y la Sección 5 concluye con conclusiones clave.

Sección 2 – Expectativas u orientaciones de las autoridades en relación con el uso de la IA por parte de las instituciones financieras

7. Como se muestra en el cuadro 1, la mayoría de las publicaciones tratadas en el presente documento adoptan la forma de documentos de debate y principios. A nivel internacional, una de las principales recomendaciones y principios a los que se hace referencia comúnmente en el sector financiero son los Principios de IA de la Organización de Cooperación y Desarrollo Económicos (OCDE)5 y los Principios de IA del G20**6**, este último a partir de las recomendaciones del primero. Estos principios abarcan cinco grandes esferas, a saber: los beneficios para las personas y el planeta; el respeto del estado de derecho y los derechos humanos; transparencia y divulgación responsable; evaluación continua de riesgos y rendición de cuentas. En esencia, estos principios requieren el uso de la IA que realmente beneficie a la sociedad en general. También se han publicado orientaciones relacionadas con la IA a nivel regional o nacional (es decir, la UE y el Reino Unido, respectivamente). Los principios que apuntan específicamente al uso de la IA en el sector financiero se han emitido en la Unión Europea, Alemania, Hong Kong, los Países Bajos, Singapur y los Estados Unidos. Más recientemente, se publicó en la UE una propuesta de Reglamento por el que se establecen normas armonizadas sobre la IA en todas las industrias, que es la primera en el mundo (véase el anexo).

8. Las emisiones existentes giran en torno a cinco principios comunes: fiabilidad/solidez, rendición de cuentas, transparencia, equidad y ética.**7** **8** El cuadro 2 resume las expectativas reglamentarias relativas a estos principios comunes. Otros temas cubiertos por las emisiones incluyen la privacidad de los datos, la dependencia de terceros y la resiliencia operativa. La mayoría de estos problemas son los mismos que las autoridades consideran al evaluar los modelos tradicionales utilizados por las instituciones financieras.**9** Lo que sigue es una discusión de estos problemas y cómo las autoridades pueden abordarlos de manera diferente en el caso de la IA en comparación con los modelos tradicionales.



9. La evaluación de la fiabilidad/solidez de los modelos de IA es similar a los modelos tradicionales, pero con énfasis en evitar daños o discriminación. La evaluación de la fiabilidad/solidez de la IA y los modelos tradicionales implica observar aspectos similares, como la validación del modelo, la definición de métricas de precisión, la actualización (o el reciclaje, en el caso de la IA) de los modelos, la garantía de la calidad de los datos utilizados por los modelos, etc. Lo que parece ser el factor distintivo es que garantizar la fiabilidad / solidez de los modelos de IA se ve desde la perspectiva de evitar causar daño (por ejemplo, discriminación) debido a decisiones inexactas que surgen del sesgo inherente en las entradas de datos o el enfoque de modelado.

10.AI cuestiones de rendición de cuentas relacionadas con el trabajo son bastante similares a las cuestiones generales de rendición de cuentas o gobernanza, pero la participación humana parece verse de manera ligeramente diferente. La rendición de cuentas se relaciona con tener roles y responsabilidades claros, así como con establecer la responsabilidad final en la junta y la alta gerencia de una institución financiera. En el caso de la IA, se hace hincapié en la intervención humana en el desarrollo de modelos y la toma de decisiones. Se enfatizan conceptos como "humano en el bucle" (intervención humana en el ciclo de decisión de la IA) y "humano en el bucle" (intervención humana durante el ciclo de diseño y revisiones posteriores). El papel del juicio humano y la revisión humana también se reconoce en las directrices existentes sobre el uso de modelos tradicionales. Lo que parece ser diferente es cómo se perciben. En el caso de los modelos tradicionales (por ejemplo, el enfoque basado en calificaciones internas para el riesgo de crédito en el marco de Basilea), las anulaciones humanas de los modelos / asignaciones de calificación deben monitorearse y documentarse para rastrear el rendimiento de las anulaciones por separado. En el caso de la IA, la intervención humana se considera más como una necesidad para garantizar que las decisiones basadas en modelos de IA no den lugar a resultados injustos o poco éticos.

11. Otro concepto de rendición de cuentas que se enfatiza cuando se trata de IA es la "responsabilidad externa". Los Principios MAS FEAT**10**, por ejemplo, establecen que las instituciones financieras que utilizan IA deben proporcionar a los interesados (por ejemplo, posibles clientes financieros) canales para preguntar, presentar apelaciones y solicitar revisiones de las decisiones impulsadas por la IA que les afectan; y tener en cuenta los datos complementarios verificados y pertinentes proporcionados por los interesados al realizar revisiones de las decisiones basadas en la IA. Si bien esto puede ser una nueva expectativa para las instituciones financieras, la práctica no es totalmente nueva. Los consumidores en los Estados Unidos, por ejemplo, tienen el derecho legal de solicitar a las agencias de informes de crédito su informe de crédito, que incluye toda la información que entra en su puntaje de crédito, y de corregirlos si hay errores.**11**

12.AI cuestiones de transparencia relacionadas con el comercio de derechos humanos se dividen en tres áreas:

 Explicabilidad**12**: hacer transparente cómo un algoritmo de IA llegó a un determinado resultado (es decir, no permitir excusas de "caja negra");

Auditabilidad: documentar el desarrollo de la IA, los procesos (incluida la toma de decisiones) y los conjuntos de datos; y

Divulgación externa: divulgación a los interesados de lo siguiente: (i) todo uso de decisiones impulsadas por la IA; ii) los datos utilizados para tomar decisiones basadas en la IA y cómo los datos afectan a la decisión; y (iii) las consecuencias de las decisiones impulsadas por la IA sobre ellos. La OCDE (2019) y el G20 (2019), en particular, señalaron que las divulgaciones deben ser en forma de información clara y fácil de entender sobre los factores y la lógica que sirvieron de base para la decisión a fin de permitir a los interesados impugnar el resultado del sistema de IA.

13. Las expectativas de explicabilidad y auditabilidad son generalmente las mismas tanto para la IA como para los modelos tradicionales e implican la divulgación interna, en particular al consejo y a la alta dirección, para que puedan comprender mejor los riesgos y las implicaciones del uso de la IA. Sin embargo, las expectativas de divulgación externa parecen ser específicas para el uso de la IA. Aunque como se mencionó anteriormente, estos son bastante similares a los requisitos legales relacionados con los informes de crédito en los Estados Unidos.

14. La equidad no es algo que normalmente se requiere explícitamente en la evaluación de los modelos tradicionales. Sin embargo, en el sector de los seguros, es común que la legislación exija a las aseguradoras que traten a los clientes de manera justa, lo que, en principio, se extiende al uso de cualquier modelo tradicional o de IA. Las cuestiones relacionadas con la equidad cubiertas en las emisiones de IA se refieren a abordar o prevenir sesgos en los algoritmos de IA que podrían conducir a resultados discriminatorios. El grupo independiente de expertos de alto nivel sobre IA creado por la Comisión Europea también se refiere a una dimensión "procedimental" de la equidad. Esto es bastante similar al concepto de responsabilidad externa discutido anteriormente. Esto implica garantizar que los sujetos de datos tengan la capacidad de impugnar y buscar una reparación efectiva contra las decisiones basadas en IA tomadas por humanos que los operan. Esto implica que la entidad responsable de la toma de decisiones debe ser identificable y el proceso de toma de decisiones explicable.

15. Sin embargo, el concepto de "equidad" es algo nebuloso.**13** Aparte de afirmar que la "equidad" puede lograrse abordando o previniendo los sesgos para no conducir a resultados discriminatorios, el concepto no está realmente definido en las emisiones. Algunos podrían argumentar que hay una diferencia entre la justicia (es decir, la equidad) y el sesgo (es decir, la predisposición), y aunque el primero debe garantizarse, hay "buenos sesgos" que deben preservarse (por ejemplo, recompensar a los conductores cuidadosos cuyos comportamientos de conducción son rastreados por la telemática con primas más bajas). Como se examina a continuación, existen leyes que tienen por objeto garantizar la equidad en la prestación de servicios financieros. Tales leyes podrían usarse para definir el concepto de equidad en el contexto de la IA, pero no son comunes en todas las jurisdicciones. Como tal, algunas jurisdicciones promueven la opinión de que las instituciones financieras individuales deben definir y poner en práctica sus propios objetivos de "equidad", similares a tener apetitos de riesgo individuales (por ejemplo, la equidad podría definirse como el cumplimiento de los valores corporativos y las normas éticas o los requisitos legales pertinentes, como la protección del consumidor).**14**

16. Algunas emisiones de IA enumeran otras formas de abordar los problemas de equidad en los modelos de IA. Las medidas van desde muy generales hasta muy concretas. Algunas de las expectativas de supervisión más concretas incluyen, por ejemplo, exigir a las empresas que establezcan un código ético de conducta para promover prácticas no discriminatorias; buscar la diversidad en los datos de entrada; revisión cuidadosa de los datos de capacitación y validación durante el proceso de capacitación del modelo; establecer políticas para la adquisición y el procesamiento legal de datos, especialmente si no están disponibles internamente; tener conjuntos de datos separados de los datos de entrenamiento y validación para verificar específicamente el sesgo del modelo; integrar normas no discriminatorias en el modelo de IA; y monitorear constantemente el rendimiento del modelo para identificar sesgos no intencionales o para asegurarse de que se comporta según lo diseñado y previsto. También hay un trabajo activo en curso en la evaluación de la equidad de los modelos de IA.**15**

17. Las cuestiones éticas son más amplias que las cuestiones de equidad. La ética implica garantizar que los clientes no serán explotados o perjudicados, ya sea a través de sesgos y discriminación, como en el caso de la equidad, o a través de otras causas (por ejemplo, IA utilizando información obtenida ilegalmente). La ética se basa en las normas o costumbres de una sociedad, que pueden codificarse en leyes, reglamentos, códigos de conducta, etc. Estos incluyen privacidad y protección de datos, no discriminación e igualdad, diversidad, inclusión y justicia social. Otra dimensión de la ética se relaciona con la cuestión de si la IA debe desplegarse en absoluto. Esto se especifica en el DNB (2019), que exige que los objetivos, estándares y requisitos para adoptar y aplicar la IA se definan en un código ético.

17. Las cuestiones éticas son más amplias que las cuestiones de equidad. La ética implica garantizar que los clientes no serán explotados o perjudicados, ya sea a través de sesgos y discriminación, como en el caso de la equidad, o a través de otras causas (por ejemplo, IA utilizando información obtenida ilegalmente). La ética se basa en las normas o costumbres de una sociedad, que pueden codificarse en leyes, reglamentos, códigos de conducta, etc. Estos incluyen privacidad y protección de datos, no discriminación e igualdad, diversidad, inclusión y justicia social. Otra dimensión de la ética se relaciona con la cuestión de si la IA debe desplegarse en absoluto. Esto se especifica en el DNB (2019), que exige que los objetivos, estándares y requisitos para adoptar y aplicar la IA se definan en un código ético.

18. A medida que los casos de uso de IA utilizan cada vez más entradas de datos de una variedad más amplia de fuentes, incluidos los datos personales, muchas emisiones regulatorias enfatizan la necesidad de cumplir con las leyes / regulaciones de privacidad y protección de datos. Estos incluyen la necesidad de garantizar que los sistemas de IA garanticen la privacidad y protección de los datos a lo largo de las diferentes etapas del proceso, así como de tener políticas que rijan el acceso a los datos y el consentimiento del cliente sobre el uso de sus datos personales. MAS (2018) afirma que el uso de atributos personales como factores de entrada para las decisiones impulsadas por la IA debe estar justificado. EIOPA (2021) destaca que los factores de calificación utilizados para la fijación de precios y la suscripción en seguros deben tener una correlación con el riesgo y un vínculo causal. El documento de discusión de CSSF (2018) va aún más lejos al recomendar la necesidad de desafiar el uso de datos personales como entrada para los modelos de IA. El documento de discusión BaFin (2018), por otro lado, señala la importancia de garantizar la libertad de elección al proporcionar productos financieros menos intensivos en datos personales.

19. Las expectativas generales de terceros o de subcontratación también son relevantes cuando se trata de la dependencia de terceros relacionada con AI. Los terceros pueden proporcionar a las empresas los datos que utilizan para sus modelos de IA o los propios modelos. Los riesgos de terceros que son relevantes en el contexto de la IA incluyen el riesgo para la privacidad y protección de los datos, la falta de comprensión de cómo funciona el modelo de IA en parte debido a las restricciones de propiedad intelectual y el riesgo de dependencia. La Autoridad de Regulación Prudencial emitió la Declaración de Supervisión SS2/21, que proporcionó ejemplos relacionados con la IA de acuerdos con terceros.**16** Las emisiones que cubren estos temas enfatizan que las políticas de IA de las instituciones financieras también deben aplicarse a terceros; que las entidades financieras cuenten con marcos de gestión de terceros, incluida la diligencia debida; y que las instituciones financieras deben ser conscientes de los riesgos de terceros. La propuesta de la UE para la regulación de la IA pretende abordar esta cuestión exigiendo a los proveedores de servicios de IA que diseñen y desarrollen sistemas de IA de una manera suficientemente transparente que permita a los usuarios interpretar la salida del sistema y utilizarla adecuadamente.**17**

20. Las cuestiones relativas a la resiliencia operativa también son pertinentes en el contexto del uso de la IA por parte de las instituciones financieras. Al igual que los modelos tradicionales, el uso de modelos de IA expone a una institución financiera a vulnerabilidades operativas. Estos incluyen fallas internas de procesos o controles, fallas de tecnología de la información, riesgos asociados con el uso de terceros, riesgo de modelo y riesgo cibernético. En términos de riesgo cibernético, los sistemas de IA pueden ser vulnerables a los ataques de "envenenamiento de datos", que intentan corromper y contaminar los datos de entrenamiento para comprometer el rendimiento del sistema.

Sección 3 – ¿Se recogen las expectativas u orientaciones relacionadas con la IA en las normas o regulaciones existentes?

21. Actualmente no existen normas reglamentarias internacionales ni orientaciones específicas sobre ia para el sector financiero. No obstante, existen normas internacionales, orientaciones y leyes nacionales que pueden aplicarse o utilizarse como puntos de partida para abordar las cuestiones de gobernanza asociadas con los modelos de IA (véase la Tabla 3 para una lista no exhaustiva de los cinco principios comunes). Esto se debe a que, como se ha mencionado, la mayoría de las cuestiones identificadas anteriormente ya existen con respecto al uso de modelos tradicionales. Esto es especialmente cierto cuando se trata de cuestiones como la fiabilidad / solidez, la responsabilidad, la transparencia, la privacidad de los datos, la dependencia de terceros, la resiliencia operativa y, hasta cierto punto, la ética. Pero este no es el caso cuando se trata de equidad, como se describe en las emisiones cubiertas en este documento (es decir, en relación con la no discriminación). Si bien existen orientaciones generales en materia de seguros para garantizar un trato justo a los clientes, este no es el caso de la banca. Sin embargo, las cuestiones de equidad están cubiertas en las leyes de protección al consumidor en algunas jurisdicciones.



22. Las normas reguladoras internacionales existentes en materia de banca y seguros pueden aplicarse en el contexto de la fiabilidad/solidez de los modelos de IA. Tanto los Principios Básicos de Basilea (BCP) como los Principios Básicos de Seguros (ICP) proporcionan orientación sobre la fiabilidad/solidez de los modelos de riesgo o los modelos utilizados con fines de capital regulatorio.**18** BcP 15, en particular sus criterios esenciales 6, establece que los bancos que utilizan modelos de riesgo deben cumplir con las normas de supervisión sobre el uso de modelos, incluida la realización de validaciones y pruebas periódicas e independientes de los modelos. La PCI 16 proporciona orientación sobre el uso de modelos para la medición del riesgo, mientras que la PCI 17 proporciona orientación sobre el uso de modelos con fines de capital regulatorio, incluida la evaluación de la idoneidad de la metodología, los insumos del modelo y las suposiciones. Los principios del BCBS para la agregación efectiva de datos de riesgo y la presentación de informes de riesgos también proporcionan orientación sobre la generación precisa y confiable de datos de riesgo por parte de los bancos.**19**

23. También existen normas que abordan la fiabilidad/solidez en ámbitos específicos en los que los modelos se utilizan con fines reglamentarios. Estos requisitos específicos del dominio (por ejemplo, el enfoque basado en calificaciones internas (IRB) para el riesgo de crédito y el enfoque de modelos internos (IMA) para el riesgo de mercado**20**; el enfoque de medición avanzada (AMA) para el riesgo operativo**21**, las pruebas de resistencia**22**; y la valoración de las provisiones técnicas**23** son bastante similares. Requieren que las instituciones financieras se aseguren de que la metodología y los supuestos del modelo sean conceptualmente sólidos, adecuados para el propósito previsto y tengan un buen poder predictivo. También requieren validación regular (por ejemplo, a través de backtesting), desafío y revisión. Además, requieren la investigación de las entradas de datos para garantizar la precisión, integridad y adecuación.

24. Existen normas de gobierno corporativo que definen los requisitos generales de rendición de cuentas para los bancos y las empresas de seguros. En particular, el BCP 14 para los bancos y el ICP 7 para los seguros. El BCBS también tiene principios dedicados al gobierno corporativo.**24** Mucho más específico para el uso del modelo es BCP 15 e ICP 17. Ambos se relacionan con el uso de modelos de riesgo y asignan la responsabilidad final a la junta y a la alta gerencia y para que comprendan las consecuencias y limitaciones de los resultados del modelo. El PCI 17 también requiere que las aseguradoras tengan una gobernanza adecuada y controles internos para los modelos de riesgo interno utilizados para determinar los requisitos de capital regulatorio.

25. Las normas específicas de dominio también requieren estructuras de gobernanza eficaces. Esto implica especificar los roles y responsabilidades para todos los aspectos del proceso de modelado. En el caso del IRB, los bancos deben tener una guía escrita que describa cómo se combinan el juicio humano y los resultados del modelo. También se requiere que los bancos cuenten con procedimientos para la revisión humana de las asignaciones de calificación basadas en modelos que se centren en encontrar y limitar los errores asociados con las debilidades conocidas del modelo y mejorar el rendimiento del modelo. Como se mencionó anteriormente, las anulaciones humanas, ya sea de la salida del modelo en sí o de la entrada utilizada, también deben identificarse y monitorearse para realizar un seguimiento del rendimiento por separado.

26. Ya existen requisitos de transparencia para los modelos utilizados con fines de supervisión. El PCI 17, por ejemplo, exige que las aseguradoras documenten los modelos internos utilizados con fines de capital regulatorio. La documentación debe abarcar el diseño del modelo, los supuestos, el cumplimiento de normativas como el uso y las pruebas estadísticas.**25** Los estándares específicos del dominio en la banca también requieren lo mismo. Además, requieren documentación para incluir políticas y procesos seguidos en el uso de los resultados del modelo, que formarán la base de las auditorías. Los principios de las pruebas de resistencia también requieren que los resultados se comuniquen dentro de las jurisdicciones y entre ellas.

27. Las leyes, normas u orientaciones reglamentarias existentes cubren suficientemente la privacidad de los datos, la dependencia de terceros y la resiliencia operativa, y pueden aplicarse en el contexto de la IA. La mayoría de las jurisdicciones tienen leyes de privacidad y protección de datos, que, como se mencionó anteriormente, ya se mencionan en las emisiones relacionadas con la IA. En el sector de los seguros, el PCI 19 exige que las aseguradoras y los intermediarios cuenten con pólizas y procesos para la protección y el uso de la información sobre los clientes. En términos de dependencia de terceros, los principios rectores del Foro Conjunto sobre la subcontratación en servicios financieros,**26** los principios del BCBS sobre resiliencia operativa y gestión del riesgo operacional**27** proporcionan orientación

28. Por último, aparte de la dependencia de terceros, los principios del BCBS sobre resiliencia operativa abarcan una amplia gama de cuestiones que pueden causar vulnerabilidades operativas, como el riesgo cibernético.**28** Las normas bancarias y de seguros existentes abarcan cuestiones relacionadas con la ética. En particular, las cuestiones éticas se mencionan en todos los PIC. En la banca, si bien solo hay un BCP que menciona la ética (BCP 29 sobre el abuso de los servicios financieros), los principios específicos del tema emitidos por el BCBS se ocupan del tema. Los requisitos éticos comienzan con el papel de la junta directiva en la definición y supervisión de la implementación del código de ética o código de conducta y garantizar que el personal reciba la capacitación ética adecuada (ICP 7, principios de gobierno corporativo de BCBS y principios para la gestión racional del riesgo operativo). El PCI 5 también requiere el establecimiento de altos estándares internos de ética e integridad para garantizar que se cumplan los requisitos para la junta, la alta gerencia y las personas clave en las funciones de control. El PCI 8 y los principios del BCBS sobre el cumplimiento y la función de cumplimiento en los bancos**29** enfatizan que el cumplimiento comienza en la parte superior. Esto significa que la junta es responsable de establecer estándares de honestidad e integridad y debe liderar con el ejemplo. Otra referencia es FSB (2018), que se centra en la mala conducta en el sector financiero, particularmente las relacionadas con la manipulación de los mercados mayoristas y los esquemas de venta indebida al por menor. Proporciona orientación general sobre el logro de un comportamiento ético por parte de las instituciones financieras, que cubriría los casos de uso de la IA.

29. Dado que la mayoría de los supervisores de seguros tienen un mandato de conducta explícito, las normas de seguros suelen cubrir cuestiones de equidad. ICP 19 requiere que las aseguradoras y los intermediarios traten a los clientes con el debido cuidado, habilidades y diligencia, y que tengan políticas y procedimientos para tratar a los clientes de manera justa. También exige que las aseguradoras y los intermediarios tengan en cuenta los intereses de los diferentes tipos de consumidores a la hora de desarrollar y distribuir productos de seguros. Además, la norma ComFrame 7.2a exige que los consejos de administración de los grupos de seguros internacionalmente activos establezcan estrategias y objetivos comerciales que tengan debidamente en cuenta el trato justo de los clientes.

30. Las normas bancarias se centran en cuestiones prudenciales y, por lo tanto, no abarcan directamente las cuestiones de equidad. El requisito del IRB de que los bancos examinen las entradas de datos para evaluar su idoneidad para asignar calificaciones puede considerarse o aplicarse en el contexto de la equidad. También se podría argumentar que las preocupaciones de equidad pueden abordarse bajo estándares de gobernanza más amplios (por ejemplo, asegurando un entorno de control sólido y la presencia de controles y equilibrios que guíen la toma de decisiones), pero hasta ahora, estas preocupaciones no han sido el foco de las normas de gobernanza para los bancos. Aparte de estas conexiones, no hay otros estándares bancarios que puedan considerarse como que abordan la equidad.

31. Las leyes de protección de los consumidores en algunas jurisdicciones abordan las preocupaciones de equidad, en particular las relacionadas con los resultados discriminatorios. Las leyes de protección del consumidor que abordan directamente las preocupaciones de equidad como se describe en las emisiones relacionadas con la IA (es decir, prevenir / abordar los sesgos para evitar resultados discriminatorios) se pueden encontrar en jurisdicciones como la UE, Sudáfrica y los EE. UU.**30** Sin embargo, la no discriminación no parece ser una característica explícita o puede no considerarse como un "área básica" de las disposiciones de trato justo en las leyes de protección al consumidor de otras jurisdicciones.**31** Los Principios de Alto Nivel sobre protección de los consumidores financieros del G20/OCDE de 2011 abarcan el trato equitativo y justo de los clientes (Principio 3), y los enfoques efectivos posteriores para apoyar la implementación de este principio incluyen un ejemplo de no discriminación.**32** Pero adoptar un enfoque muy amplio de la discriminación podría ser difícil y la industria parece preferir no abordar el problema explícitamente.**33**

Sección 4 – Desafíos en la implementación de las expectativas u orientaciones relacionadas con la IA

32. Al igual que con cualquier marco regulatorio, las instituciones financieras deben implementar los principios, requisitos u orientaciones de gobernanza de la IA para alcanzar los objetivos políticos previstos. Si bien los principios de alto nivel y la orientación sobre la gobernanza de la IA por parte de las organizaciones internacionales y las autoridades nacionales son bienvenidos, las instituciones financieras piden una orientación más detallada que ilustre cómo se pueden implementar esos principios en la práctica. Según las autoridades tratadas en este documento, hay poca necesidad de normas más generales o de alto nivel sobre la gobernanza de la IA, ya que los principios comunes descritos en la Sección 2 son bien entendidos y aceptados. En los párrafos siguientes se esbozan los problemas concretos de aplicación relacionados con cada uno de esos principios comunes.

Transparencia

33. Las expectativas supervisoras sobre la transparencia de los modelos utilizados por las entidades financieras y de que tales modelos no deben ser «cajas negras» no son nuevas. La principal diferencia entre la IA y los modelos más tradicionales es el nivel de complejidad y la falta de explicabilidad de ciertos tipos de algoritmos de ML. Explicar algoritmos complejos de ML de una manera que pueda ser entendida por un supervisor puede ser un desafío. Los supervisores necesitan mejorar las habilidades de su personal para estar familiarizados con las técnicas de ML. Las empresas deben hacer más esfuerzos para explicar sus modelos de ML de una manera comprensible. Una prueba de fuego práctica es evaluar la comprensión del concepto general del modelo por parte de los miembros de su junta; si esto puede ser entendido por los miembros no técnicos de la junta, lo más probable es que los supervisores también los entiendan. Algunos algoritmos de ML, como el aprendizaje profundo y las redes neuronales, se consideran como "caja negra"**34**, que producen resultados sensatos, pero son difíciles de explicar o probar a diferencia de los modelos estadísticos tradicionales. Esto se debe a que tales algoritmos de ML funcionan a través de interacciones complejas entre múltiples variables, infiriendo atributos a partir de entradas de datos y colocando diferentes pesos en diferentes atributos de datos. Esto también hace que sea difícil revelar a los interesados en un lenguaje sencillo y sencillo qué datos se utilizan y cómo afectan a la decisión relevante para ellos.

34. La transparencia de un algoritmo de IA/ML es un requisito previo para cumplir algunos de los otros principios sólidos de gobernanza de la IA. Si un modelo no es transparente, será difícil evaluar su fiabilidad, rendimiento y equidad, aparte de evaluar el resultado del modelo con respecto a un punto de referencia específico. También será difícil establecer la responsabilidad si no está claro qué componentes del algoritmo están causando errores. CSSF (2018) explicó que esto es particularmente importante cuando se utilizan soluciones listas para usar que automatizan la toma de decisiones porque pueden surgir disputas cuando se toman decisiones incorrectas.

35. Las entidades financieras deben aplicar un buen juicio al determinar el nivel adecuado de transparencia de sus modelos de IA/ML en función del público objetivo y la materialidad de los resultados de los modelos. CSSF (2018) destacó que cuanto más crítico es el caso de uso, más importante debe ser un algoritmo para ser transparente. De lo contrario, el usuario empresarial podría refutar los resultados del modelo y las personas responsables podrían perder la confianza en el modelo y negarse a asumir la responsabilidad. La transparencia inadecuada también podría erosionar la confianza del consumidor o disuadir a los clientes de usar soluciones financieras impulsadas por IA / ML. Por otro lado, MAS (2018) señaló que la transparencia excesiva podría crear confusión u oportunidades no deseadas para que las personas, incluidos los clientes de instituciones financieras, exploten o manipulen los modelos de IA / ML EIOPA (2021) considera que las aseguradoras deben esforzarse por utilizar algoritmos explicables, en particular para aplicaciones de IA de alto impacto, incluso si esto es a expensas del rendimiento del modelo. En algunos casos en los que no hay alternativas viables (por ejemplo, el procesamiento de imágenes, videos o textos), las aseguradoras pueden utilizar medidas de gobernanza alternativas, como una mayor supervisión humana.

Fiabilidad y solidez

36. Con respecto a la fiabilidad y solidez de los resultados de la IA, el principal desafío de implementación surge de la construcción técnica de los algoritmos de IA. Una forma de enmarcar este problema es mediante el uso de la canalización de ciencia de datos. En cada paso de la tubería de ciencia de datos, se pueden implementar salvaguardas técnicas para facilitar resultados confiables y sólidos. A continuación, se presentan ejemplos de desafíos prácticos que pueden surgir en cada paso de la canalización de la ciencia de datos: Recopilación de datos: conjuntos de datos de origen inexactos, incompletos o sesgados conducirán a resultados de modelado poco confiables o sesgados; la fuente de datos no podrá generalizarse para los grupos de consumidores destinatarios previstos. Limpieza y procesamiento de datos: esta suele ser la parte que más tiempo consume de la implementación de un algoritmo de IA: los recursos insuficientes asignados a este paso o la falta de rigor en la limpieza de los datos conducirán a los proverbiales resultados de "basura en la basura". El sesgo presente en los datos de origen, que puede ser difícil de detectar durante la limpieza de datos, o introducido a través de la intervención humana puede ser amplificado por los algoritmos de IA. Análisis de datos: no es difícil determinar qué datos deben usarse como conjunto de entrenamiento (para entrenar el algoritmo) y cuáles usar como conjunto de pruebas (para verificar los resultados).

37. Dado que las técnicas de aprendizaje automático, por definición, aprenden de nuevos datos a lo largo del tiempo, otro desafío de implementación para mantener resultados confiables y sólidos se relaciona con la actualización y revisión periódicas y oportunas de los resultados del modelo. El Banco de Inglaterra y la Autoridad de Conducta Financiera (2020) destacaron que las fuentes de datos subyacentes de un modelo de IA o sus propiedades estadísticas pueden cambiar con el tiempo, por ejemplo, debido a los cambios en el comportamiento del consumidor derivados de la aceleración digital provocada por Covid-19. El ML supervisado requiere una retroalimentación humana constante sobre el rendimiento del algoritmo para que pueda aplicar su "conocimiento" entrenado a nuevos datos y continuar produciendo resultados confiables. Si el contexto cambia, el conjunto de datos de entrenamiento deberá actualizarse para que el algoritmo pueda volver a entrenarse para producir resultados confiables**35**. Incluso el ML no supervisado que aprende y produce resultados por sí mismo requiere actualizaciones y revisiones oportunas. Los patrones o relaciones reveladas deben ser evaluados por un experto humano para determinar que tienen sentido.

38. Es posible que sea necesario ajustar los requisitos reglamentarios existentes sobre la validación de modelos para adaptarse a determinadas técnicas de aprendizaje automático. Específicamente, los cambios significativos en los modelos internos utilizados por las instituciones financieras generalmente requieren una nueva aprobación supervisora. Lo que constituye "cambios" en el contexto de las técnicas de ML puede necesitar ser redefinido. Por ejemplo, el ML supervisado ajusta el modelo a medida que aprende de los nuevos datos. Sin embargo, no tiene sentido definir cada ajuste como un "cambio" en el modelo que requiere la aprobación de la supervisión, a menos que el rendimiento o los resultados del modelo cambien significativamente.

39. Puede ser difícil lograr el equilibrio adecuado entre la simplicidad de un algoritmo y su rendimiento. Por ejemplo, la "precisión" es una de las medidas de rendimiento que evalúa la proporción de resultados correctamente identificados sobre todos los resultados de un algoritmo. Para mejorar la precisión, una empresa podría agregar más elementos, pero potencialmente a expensas de hacer que el modelo sea más complejo y, por lo tanto, más difícil de explicar y divulgar.

40. Es posible que no se puedan alcanzar requisitos reglamentarios excesivamente onerosos para «probar» la exactitud de un algoritmo de ml. ACPR (2020b) destacó tal desafío al cumplir con el criterio de explicabilidad más estricto propuesto que busca demostrar que un algoritmo funciona correctamente. En la práctica, esto requiere una revisión línea por línea del código fuente, un análisis exhaustivo de todos los conjuntos de datos utilizados y un examen del modelo y sus parámetros, que algunas empresas consideran inalcanzables. EIOPA (2021) especifica los principios de gobernanza de la IA que adaptan la intensidad de las medidas de gobernanza al impacto de un caso de uso de IA determinado.

41. La creación y ejecución de un algoritmo de aprendizaje automático requiere un ejercicio significativo de juicio técnico y experiencia, cada elección de los cuales podría producir resultados muy diferentes. Esto subraya la importancia de la experiencia técnica humana en el desarrollo, ejecución y validación de un modelo de IA para que produzca resultados confiables y sólidos. La falta de experiencia técnica en las empresas podría limitar la medida en que pueden utilizar las tecnologías de IA. La selección de medidas de rendimiento adecuadas es un ejemplo en el que se necesita un buen juicio. LA AESPJ (2021) proporciona un ejemplo en la detección de fraudes, mediante el cual las aseguradoras deben decidir si el objetivo es maximizar la precisión de la predicción (número de siniestros fraudulentos detectados), reducir el número de falsos positivos (créditos legítimos etiquetados erróneamente como fraudulentos) o falsos negativos (siniestros etiquetados como legítimos que al final son fraudulentos).

42. El mayor uso de la IA/ML en los procesos empresariales por parte de las instituciones financieras debe estar respaldado por un marco sólido de ciber-resiliencia. CSSF (2018) explica cómo una empresa puede estar expuesta a ataques de envenenamiento de datos que implican alterar el conjunto de datos de entrenamiento utilizado para construir un algoritmo predictivo de ML con el fin de manipular sus resultados. El ataque adversario es otra forma de ataque cibernético mediante el cual un adversario proporciona conjuntos de datos alterados para que el modelo entrenado clasifique erróneamente los datos. Tales violaciones cibernéticas podrían conducir a resultados de ML poco confiables y potencialmente volverse sistémicos si no se detectan o amplifican a través de la escalabilidad del modelo.

Responsabilidad

43. Existen importantes desafíos para cumplir con el criterio de rendición de cuentas de la implementación de la IA, entre otras cosas porque esto a menudo conduce a la reflexión filosófica de la demarcación entre máquinas y humanos. En general, los supervisores suelen colocar la responsabilidad final de cualquier decisión o acción de una institución financiera en su junta directiva y alta gerencia, y con razón. Si bien esto debería aplicarse de manera similar a la implementación de ia dentro de una empresa, la responsabilidad se vuelve menos clara en los niveles inferiores de la jerarquía. Por ejemplo, si un algoritmo de IA niega sistemáticamente las solicitudes de préstamo de solicitantes negros que viven dentro de un determinado código postal, ¿debería ser responsable el desarrollador de IA, o el jefe del departamento de crédito o, de hecho, el CEO?

44. Las salvaguardias "human-in-the-loop" o human-on-the-loop podrían dar lugar a nuevos riesgos relacionados con la gobernanza. ACPR (2020a) explica que un operador humano que invalida los resultados correctamente identificados de una máquina podría ser considerado responsable del error. También podría haber una tendencia de un operador humano a aceptar los resultados de un modelo para evitar tener que justificar cualquier desviación de puntos de vista. La intervención humana también podría introducir sesgos y complicar la transparencia de un algoritmo, teniendo que explicar anulaciones humanas subjetivas. En términos más generales, la intervención humana en el diseño y las decisiones de los modelos de IA requiere el reclutamiento, la capacitación y el reciclaje del personal con experiencia especializada, así como la mejora de las habilidades de las juntas directivas y la alta gerencia con responsabilidad explícita por el despliegue de la IA. Esas necesidades de recursos humanos podrían plantear un reto importante a las instituciones financieras.

45. AI/ML implementación a menudo requiere la subcontratación por parte de las instituciones financieras a proveedores de servicios externos, lo que plantea desafíos únicos en la implementación de marcos de gobernanza de IA sólidos. Las interacciones con terceros proveedores de servicios a través de la concesión de licencias de sistemas de IA y la adquisición de datos (por ejemplo, puntajes de crédito) podrían imponer restricciones a las instituciones financieras derivadas de los derechos de propiedad intelectual impuestos por dichos proveedores. Crisanto et al (2018) encontraron que la mayoría de las autoridades reguladoras financieras esperan que sus instituciones reguladas conserven la plena responsabilidad y rendición de cuentas de los servicios subcontratados. Dichos requisitos deben establecer claridad en cuanto a la parte responsable de los resultados deficientes de IA/ML. La dependencia excesiva de proveedores de servicios externos también podría conducir a la captura comercial y al riesgo de dependencia, particularmente dada la creciente venta cruzada de servicios digitales por parte de las bigtechs, lo que dificulta que una institución financiera contrate las habilidades y la experiencia cuando sea necesario, por ejemplo, en un escenario de continuidad del negocio.

Equidad y ética

46. Un reto fundamental en la aplicación de las orientaciones o principios "éticos" y "éticos" es la falta de definiciones universalmente aceptadas de estos términos. Si bien es apropiado que algunos supervisores hayan dejado en manos de las empresas la elaboración de sus propias definiciones, algunas empresas pueden tener dificultades para hacerlo. Incluso si pudieran, las definiciones podrían no estar a la altura de las expectativas generales de los consumidores. Otros reguladores, como la AESPJ (2021), proporcionan más orientación, por ejemplo, la definición de «uso leal de los datos», que define como la garantía de que los datos son adecuados para su finalidad y respeta el principio de autonomía humana mediante el desarrollo de sistemas de IA que apoyen a los consumidores en su proceso de toma de decisiones. La importancia de la equidad y la ética no debe subestimarse, especialmente dados los fuertes movimientos recientes de justicia social como Black Lives Matter. Una reacción pública no solo podría infligir daños a la reputación de una institución financiera, sino que también podría causar problemas de solvencia o liquidez si los clientes cambian de proveedor en masa. Para los supervisores prudenciales, tratar con la equidad y la ética es particularmente desafiante, ya que generalmente no están dentro de su competencia ni experiencia.

47. Las regulaciones que requieren el ejercicio de un buen juicio humano pueden dificultar la implementación de AI / ML. ACPR (2020a) citó un ejemplo de seguros, refiriéndose a la directiva europea IDD (Directiva de Distribución de Seguros) de 2016 que requiere que los distribuidores de productos de seguros "actúen siempre de manera honesta, justa y profesional de acuerdo con los mejores intereses de sus clientes". Puede ser difícil cumplir con este requisito cuando se utiliza ML para identificar las necesidades futuras de seguros de un consumidor, ya que la consideración de la información contextual probablemente requerirá una evaluación humana.

48. La aplicación injusta o poco ética de la IA podría tener profundas consecuencias para los consumidores, que incluyen la exclusión financiera con la consiguiente destrucción de los medios de subsistencia. CSSF (2018) describió varias fuentes de sesgo: sesgo algorítmico que surge de la elección de un modelo incorrecto o sesgo humano que se refleja en un conjunto de datos de entrenamiento. El documento proporcionó un ejemplo de calificación crediticia, por el cual es poco probable que las poblaciones que no están bien representadas en el conjunto de datos reciban una puntuación de crédito favorable simplemente porque el algoritmo ha aprendido que, en el pasado, a dichos solicitantes no se les otorgaron muchos préstamos. Esto plantea un desafío comercial para las empresas, ya que pueden estar excluyendo inadvertidamente a las empresas rentables. BaFin (2018) destacó cómo los consumidores ni siquiera son conscientes de que están siendo discriminados, lo que resulta en la denegación de acceso a ciertos servicios financieros. Tal situación puede ocurrir a través del poder predictivo de AI / ML de futuros cambios en el estilo de vida al inferir de los cambios en los patrones de gasto, incluso sin recopilar ningún dato específico sobre esos cambios.

49. En general, los principios de gobernanza de la IA o la orientación sobre equidad y ética requieren que el ser humano en el bucle o el ser humano en el bucle. En la práctica, esto significa que las instituciones financieras deben asignar suficientes recursos humanos en cualquier implementación de IA para determinar resultados justos y éticos. En cierto modo, hay cierta ironía en tener a los humanos como salvaguardas para la injusticia y los resultados poco éticos de la IA cuando este último básicamente solo refleja estos defectos humanos. En cualquier caso, dado que uno de los principales beneficios de la IA es reducir la necesidad de intervención humana, será un desafío lograr un equilibrio adecuado en el cumplimiento de las expectativas humanas en el entorno frente a la cosecha de todos los beneficios de la automatización de la tecnología.

Abordar los desafíos regulatorios/supervisores a través de la proporcionalidad

50. Debido a los desafíos señalados anteriormente, las autoridades están estudiando la posibilidad de aplicar la proporcionalidad en la aplicación de los principios comunes. Esto requiere diferenciar el tratamiento regulatorio y supervisor de los modelos de IA/ML en función de la conducta y los riesgos prudenciales que plantean. Por ejemplo, los modelos de IA utilizados para las decisiones de suscripción de crédito podrían estar sujetos a una mayor equidad y expectativas éticas que los utilizados para los chatbots de atención al cliente. Del mismo modo, se podría esperar un mayor nivel de fiabilidad/solidez de los modelos de IA utilizados para el cálculo del capital regulatorio en relación con los utilizados para los procesos operativos internos. El recuadro 1 presenta una posible forma de avanzar hacia la aplicación de este enfoque regulador y de supervisión adaptado.

51. Ya hay ejemplos de enfoques adaptados similares. El Reglamento de IA propuesto en la UE identifica actividades prohibidas (por ejemplo, el uso de IA en la administración de justicia y procesos democráticos) y sistemas de IA de alto riesgo (por ejemplo, sistemas de IA que evalúan la solvencia de las personas físicas o establecen su puntaje de crédito). La AESPJ (2021) ha desarrollado un marco de evaluación del impacto de los casos de uso de ia para ayudar a las empresas de seguros a determinar de manera proporcionada la combinación de medidas de gobernanza necesarias para un caso de uso concreto de IA. Efectivamente, las actividades de mayor impacto atraerían medidas de gobernanza más integrales, y viceversa. Algunas autoridades también están pensando en exigir una estrategia general de IA aprobada por la junta si el despliegue de IA de una empresa excede ciertos umbrales. Dicha estrategia debe estar conectada con las políticas pertinentes, por ejemplo, códigos de conducta, resiliencia operativa, gestión de terceros, uso de modelos, etc. Como tal, el amplio despliegue de la IA en las principales líneas de negocio de las empresas (por ejemplo, la suscripción de crédito o seguros) podría requerir una estrategia de junta de este tipo.



Sección 5 – Conclusión

52. La mayoría de las cuestiones relacionadas con el uso de la IA por parte de las instituciones financieras son bastante similares a las de los modelos tradicionales, pero la perspectiva podría ser diferente. Entre los problemas comunes identificados por las emisiones relacionadas con la IA cubiertas en este documento, la confiabilidad / solidez, la responsabilidad, la transparencia, la privacidad de los datos, la dependencia de terceros y la resiliencia operativa son relevantes en el uso de la IA y los modelos tradicionales. Solo las cuestiones relacionadas con la equidad son únicas o explícitas para la IA. Sin embargo, cuando se trata del uso de la IA, algunos de los problemas comunes identificados anteriormente se ven desde la perspectiva de la equidad. Por ejemplo, garantizar la fiabilidad/ solidez de los modelos de IA tiene como objetivo evitar causar discriminación debido a decisiones inexactas. Además, garantizar la rendición de cuentas y la transparencia en el uso de la IA incluye asegurarse de que los interesados conozcan las decisiones basadas en datos, los datos utilizados y cómo afectaron a las decisiones, y tengan canales para preguntar y desafiar estas decisiones.

53. Si bien las normas, leyes y orientaciones existentes pueden utilizarse para abordar la mayoría de las cuestiones relacionadas con la IA, puede haber margen para hacer más cuando se trata de equidad. Dado que la mayoría de los problemas asociados con el uso de la IA son similares a los de los modelos tradicionales, las autoridades pueden aprovechar las normas, leyes y orientaciones existentes destinadas a estos últimos al evaluar los primeros. La equidad, en particular en lo que se refiere a evitar resultados discriminatorios, si bien se identifica como muy importante en el caso de la IA, puede no ser explícita en las leyes de protección del consumidor en algunas jurisdicciones. Hacer explícitos los objetivos de no discriminación puede ayudar a proporcionar una buena base para definir la equidad en el contexto de la IA, proporcionar una base legal para que las autoridades financieras emitan orientaciones relacionadas con la IA y, al mismo tiempo, garantizar que las decisiones impulsadas por la IA, basadas en modelos tradicionales y humanas en los servicios financieros se evalúen con arreglo al mismo estándar.

54. La mayoría de los supervisores se encuentran todavía en las primeras etapas del desarrollo de principios de gobernanza específicos de la IA u orientación para las empresas financieras. Si bien esta orientación de alto nivel puede ser útil para resistir la prueba del tiempo y ser lo suficientemente duradera como para aplicarla de una manera independiente de la tecnología (un rasgo clave en el mundo de la tecnología en rápido desarrollo), a las empresas les resultaría útil contar con una orientación práctica más concreta. La mayoría de los supervisores están trabajando en asociación con la industria y otras partes interesadas en la tecnología para desarrollar dicha orientación.

55. A falta de orientaciones prácticas concretas o de expectativas de supervisión sobre la gobernanza de la IA, a algunas empresas les resulta difícil establecer adecuadamente salvaguardias en sus implementaciones de IA. Los desafíos que enfrentan las empresas son muy variados, desde dificultades técnicas en la construcción y ejecución de modelos de ML que ofrecen resultados confiables incluso en circunstancias cambiantes, hasta problemas sociales más amplios que surgen de resultados injustos o poco éticos. El primer paso para abordar estos desafíos es esforzarse por lograr modelos transparentes y explicables. Este es un requisito previo para diagnosticar los problemas que surgen de la confiabilidad / solidez de los resultados, la responsabilidad y la equidad y la ética.

56. Los retos y la complejidad que plantea la IA exigen una respuesta normativa y supervisora adaptada y coordinada basada en las implicaciones del modelo de IA para la conducta y los riesgos prudenciales. Cuanto más pueda afectar el uso del modelo de IA a la conducta y los objetivos prudenciales de las autoridades, más estrictos serán los requisitos pertinentes de fiabilidad/solidez, rendición de cuentas, transparencia, equidad y ética. Además, el uso de la IA por parte de las instituciones financieras tendrá implicaciones para la rentabilidad, el impacto en el mercado, la protección del consumidor y la reputación. Esto requiere una mayor coordinación entre las autoridades prudenciales y de conducta en la supervisión del despliegue de la IA en los servicios financieros.

57. Habida cuenta de los nuevos temas comunes sobre la gobernanza de la IA en el sector financiero, parece haber margen para que los organismos de normalización financiera elaboren orientaciones o normas internacionales en este ámbito. Las opiniones de las autoridades sobre la forma en que deben aplicarse estos temas comunes siguen evolucionando. Un intercambio continuo de opiniones y experiencias a nivel internacional podría conducir eventualmente a la elaboración de normas internacionales. Tales estándares internacionales podrían ser útiles, particularmente para las jurisdicciones que recién están comenzando su viaje de transformación digital. También pueden servir como un punto de referencia mínimo para guiar el despliegue ordenado de las tecnologías de IA dentro del sector financiero. A medida que surjan enfoques regulatorios más específicos o expectativas de supervisión sobre aspectos específicos de los casos de uso de IA, los organismos normativos podrían estar en una mejor posición para identificar las mejores prácticas comunes que serán útiles para que otras jurisdicciones las consideren.

Anexo – Propuesta de Reglamento sobre IA en la UE

El reglamento propuesto define el "sistema de inteligencia artificial" (sistema de IA) como un software que se desarrolla con una o más de las siguientes técnicas y enfoques que pueden, para un conjunto determinado de objetivos definidos por humanos, generar resultados como contenido, predicciones, recomendaciones o decisiones que influyen en los entornos con los que interactúan:

a) Enfoques de aprendizaje automático, incluido el aprendizaje supervisado, no supervisado y de refuerzo, utilizando una amplia variedad de métodos, incluido el aprendizaje profundo.

b) Enfoques basados en la lógica y el conocimiento, incluida la representación del conocimiento, la programación inductiva (lógica), las bases de conocimiento, los motores de inferencia y deductivos, el razonamiento (simbólico) y los sistemas expertos

c) Enfoques estadísticos, métodos bayesianos de estimación, búsqueda y optimización.

El Reglamento propuesto identifica las prácticas prohibidas, que comprenden todos aquellos sistemas de IA cuyo uso se considera contrario a los valores de la UE. Estos incluyen sistemas de IA que manipulan a las personas a través de técnicas subliminales más allá de su conciencia, dan como resultado una puntuación social para fines generales realizada por las autoridades públicas y utilizan sistemas de identificación biométrica remota en espacios de acceso público con el propósito de hacer cumplir la ley, excepto en ciertas situaciones limitadas.

El Reglamento propuesto también identifica los sistemas de IA de alto riesgo como los utilizados en los siguientes ámbitos:

a) Identificación biométrica y categorización de personas físicas

b) Gestión y operación de infraestructura crítica

c) Educación y formación profesional

d) Empleo, gestión de los trabajadores y acceso al trabajo por cuenta propia;

e) Acceso y disfrute de servicios privados esenciales y servicios y beneficios públicos (aquí se incluyen sistemas de IA que evalúan la solvencia de las personas físicas o establecen su puntaje de crédito

f) Aplicación de la ley

g) Gestión de la migración, el asilo y el control de las fronteras

h) Administración de justicia y procesos democráticos.

El Reglamento propuesto impone los siguientes requisitos para los sistemas de IA de alto riesgo:

a) Sistema de gestión de riesgos: para las entidades de crédito reguladas en virtud de la Directiva sobre requisitos de capital, los requisitos de gestión de riesgos para los sistemas de IA de alto riesgo pertinentes formarán parte de los procedimientos de gestión de riesgos establecidos de conformidad con dicha Directiva.

b) Datos y gobernanza de datos: los conjuntos de datos de formación, validación y ensayo estarán sujetos a prácticas adecuadas de gobernanza y gestión de datos

c) Documentación técnica: la documentación debe demostrar que los sistemas de IA de alto riesgo cumplen con todos los requisitos del reglamento y proporcionar a las autoridades toda la información necesaria para evaluar el cumplimiento de dichos requisitos.

d) Mantenimiento de registros: los sistemas de IA de alto riesgo se diseñarán y desarrollarán con capacidades que permitan el registro automático de eventos ("registros") mientras estén en funcionamiento.

e) Transparencia y suministro de información a los usuarios: los sistemas de IA de alto riesgo irán acompañados de instrucciones que incluyan información concisa, completa, correcta y clara que sea relevante, accesible y comprensible para los usuarios.

f) Supervisión humana: la supervisión humana tendrá por objeto prevenir o minimizar los riesgos para la salud, la seguridad o los derechos fundamentales que puedan surgir al utilizar un sistema de IA de alto riesgo.

g) Precisión, robustez y ciberseguridad: los sistemas de IA de alto riesgo serán resistentes en lo que respecta a los errores, fallos o incoherencias que puedan producirse en el sistema o en el entorno en el que opera el sistema, en particular debido a su interacción con personas físicas u otros sistemas; la solidez de los sistemas de IA de alto riesgo puede lograrse mediante soluciones de redundancia técnica; las soluciones técnicas destinadas a garantizar la ciberseguridad de los sistemas de IA de alto riesgo serán adecuadas a las circunstancias y los riesgos pertinentes.

1. Jermy Prenio (Jermy.Prenio@bis.org) y Jeffery Yong (Jeffery.Yong@bis.org), Banco de Pagos Internacionales. Agradecemos a Douglas Araujo, Julián Arévalo, Orlando Fernández Ruiz, Denise García Ocampo, Xuchun Li y Oliver Thew por sus útiles comentarios. Luciana D'Agnone proporcionó un valioso apoyo administrativo.

2. Para un micro y macroanálisis de los posibles efectos de la adopción de la IA en los mercados financieros, véase FSB (2017).

3. Véase, por ejemplo, EIU (2020).

4. La Autoridad Bancaria Europea, la Autoridad Europea de Seguros y Pensiones de Jubilación, la Autoridad Monetaria de Singapur y la Autoridad de Regulación Prudencial del Reino Unido

5. Véase OCDE (2019).

6. Véase G20 (2019).

7. Se reconoce que las diferentes autoridades pueden utilizar otros términos para caracterizar conceptos similares o pueden agrupar determinados conceptos (por ejemplo, fiabilidad/solidez en el marco de la equidad). La forma en que el artículo llama o distingue los diferentes conceptos se basa en el juicio de los autores.

8. Véase Fjeld et al (2020) para una revisión más extensa de los principios relacionados con la IA, incluidos los sectores privado y no financiero.

9. A los efectos de este documento, los "modelos tradicionales" se refieren a los modelos utilizados por las instituciones financieras que no utilizan algoritmos de IA o ML y para los que las autoridades han examinado o emitido directrices. Estos incluyen, por ejemplo, modelos utilizados para la evaluación de riesgos o cálculos de capital regulatorio.

10. Ver MÁS (2018).

11. Véase [www.myFICO.com](http://www.myFICO.com).

12. Existe una discusión académica en curso sobre la diferencia entre la "explicabilidad" como capaz de explicar las cajas negras (es decir, la explicación es un modelo separado que se supone que replica la mayoría de los comportamientos de una caja negra) y la "interpretabilidad", lo que significa diseñar modelos de IA que sean inherentemente interpretables (ver Rudin, 2019).

13. Hay más de 20 definiciones matemáticas de equidad (véase, por ejemplo, Verma y Rubin (2018)).

14. Véase Veritas Consortium (2020).

15. Ibíd.

16. Estos incluyen la compra de datos recopilados por proveedores externos (corredores de datos), por ejemplo, datos geoespaciales o datos de la actividad del dispositivo en la aplicación, redes sociales, etc.; y modelos de ML "listos para usar" de software de código abierto y bibliotecas de ML desarrolladas por proveedores externos.

17. Artículo 13 de la propuesta de Reglamento de la UE sobre ia.

18. Véanse BCBS (2019) e IAIS (2019), respectivamente.

19. Véase BCBS (2013).

20. Véase BCBS (2019).

21. Aunque el uso de AMA se eliminará como una opción para calcular el capital regulatorio en el marco de Basilea para 2023.

22. Véase BCBS (2018).

23. Véase IAIS (2019), en particular la Norma 16.7.d del Marco Común para la Supervisión de Grupos de Seguros Internacionalmente Activos (ComFrame).

24. Véase BCBS (2015).

25. Una "prueba de uso" es el proceso mediante el cual se evalúa un modelo interno en términos de su aplicación dentro de los procesos de gestión de riesgos y gobernanza de una aseguradora. Una "prueba de calidad estadística" evalúa la metodología cuantitativa de un modelo interno.

26. El Foro Conjunto (2005).

27. Véase BCBS (2021a) y (2021b), respectivamente.

28. Véase FSB (2020).

29. Véase BCBS (2005).

30. La Directiva de la Unión sobre créditos al consumo respeta los principios reconocidos, en particular, por la Carta de los Derechos Fundamentales de la Unión, en particular la no discriminación por motivos tales como el sexo, la raza, el color, el origen étnico o social, las características genéticas, la lengua, la religión o las convicciones, etc. La Ley de protección del consumidor de Sudáfrica también prohíbe la discriminación tal como se define en su Constitución o en la Ley de promoción de la igualdad y prevención de la discriminación injusta. En los Estados Unidos, la Ley de Vivienda Justa (FHA) prohíbe la discriminación en las transacciones relacionadas con bienes raíces residenciales basadas en: raza o color; origen nacional; religión; sexo; estado familiar; y hándicap. La Ley de Igualdad de Oportunidades de Crédito (ECOA) prohíbe la discriminación en las transacciones de crédito basadas en: raza o color; origen nacional; religión; sexo; estado civil; edad (siempre que el solicitante tenga la capacidad de celebrar un contrato); la recepción del solicitante de ingresos de un programa de asistencia pública; y el ejercicio del solicitante, de buena fe; de cualquier derecho en virtud de la Ley de Protección del Crédito al Consumidor.

31. Ardic et al (2011) centraron su encuesta en las leyes y reglamentos de protección del consumidor en lo que considera como las cuatro áreas básicas de las disposiciones de trato justo en las leyes de protección del consumidor, que son: (1) restricciones a la publicidad engañosa; (2) cobros abusivos; (3) prácticas de venta desleales o de alta presión; y (4) violación de la confidencialidad del cliente.

32. G20 y OCDE (2011) y (2019), respectivamente.

33. IC (2013).

34. Véase CRO Forum (2019).

35. ACPR (2020a) propone la estabilidad como uno de los cuatro principios para medir los algoritmos de IA. La deriva temporal derivada de los cambios en la distribución de probabilidad de los datos de entrada podría socavar el rendimiento de un algoritmo de IA si no se vuelve a entrenar periódicamente utilizando datos actualizados.

Referencias

Ardic, O, J Ibrahim y N Mylenko (2011): "Leyes y regulaciones de protección al consumidor en servicios de depósito y préstamo: un análisis entre países con un nuevo conjunto de datos", enero.

Banco de Inglaterra y Autoridad de Conducta Financiera (2019): "Machine Learning in UK financial services", octubre.

——— (2020): “Minutes: artificial intelligence public-private forum – first meeting”, October.

Comité de Supervisión Bancaria de Basilea (2005): "Compliance and the compliance function in banks", abril.

——— (2013): “Principles for effective risk data aggregation and risk reporting”, January.

——— (2015): “Corporate governance principles for banks”, July.

——— (2018): “Stress testing principles”, October.

——— (2019): “Consolidated Basel framework”, December.

——— (2021a): “Principles for operational resilience”, March.

——— (2021b): “Revisions to the principles for the sound management of operational risk”, March

Consumers International (2013): "En busca de buenas prácticas en la protección del consumidor financiero", febrero.

Crisanto, J C, C Donaldson, D Garcia-Ocampo y J Prenio (2018): "Regulating and supervising the clouds: emerging prudential approaches for insurance companies", FSI Insights on policy implementation, no 13, diciembre.

CRO Forum (2019): "Decisiones de máquina: gobernanza de la IA y análisis de big data", mayo.

Economist Intelligence Unit (2020): "El camino a seguir: la inteligencia artificial y el futuro de los servicios financieros".

Autoridad Bancaria Europea (2020): "Informe sobre big data y analítica avanzada", enero.

Comisión Europea (2021): «Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas sobre inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión», abril.

Autoridad Europea de Seguros y Pensiones de Jubilación (2019): "Big data analytics in motor and health insurance: a thematic review", mayo.

——— (2021): “Artificial intelligence governance principles: towards ethical and trustworthy artificial intelligence in the European insurance sector”, June.

Consejo de Estabilidad Financiera (2017): "Inteligencia artificial y aprendizaje automático en los servicios financieros: desarrollos del mercado e implicaciones para la estabilidad financiera", noviembre.

——— (2018): “Strengthening governance frameworks to mitigate misconduct risk: a toolkit for firms and supervisors”, April.

——— (2020): “Regulatory and supervisory issues relating to outsourcing and third-party relationships”, November.

Fjeld J, N Achten, H Hilligoss, A C Nagy y M Srikumar (2020): "Inteligencia artificial basada en principios: mapeo del consenso en enfoques éticos y basados en derechos para los principios de la IA".

Autoridad Francesa de Supervisión y Resolución Prudencial (ACPR) (2020a): "Governance of artificial intelligence in finance", junio.

——— (2020b): “Governance of artificial intelligence in finance – summary of consultation responses”, December.

G20 (2019): "Declaración ministerial del G20 sobre comercio y economía digital", junio.

G20 y Organización de Cooperación y Desarrollo Económicos (2011): "High-level principles on financial consumer protection", octubre.

——— (2019): “Compendium of effective approaches for financial consumer protection in the digital age: FCP principles 1, 2, 3, 4, 6, 7, 8 and 9”.

Autoridad Federal Alemana de Supervisión Financiera (BaFin) (2018): "Big data meets artificial intelligence", julio.

——— (2021): “Big data and artificial intelligence: principles for the use of algorithms in decision-making processes”, June.

Autoridad Monetaria de Hong Kong (2019a): "Protección del consumidor con respecto al uso de análisis de big data e inteligencia artificial por parte de instituciones autorizadas", noviembre.

——— (2019b): “High-level principles on AI”, November.

Grupo de Expertos Independientes de Alto Nivel sobre Inteligencia Artificial (2019): "Ethics guidelines for trustworthy AI", abril.

Asociación Internacional de Supervisores de Seguros (2019): "Insurance core principles and common framework for the supervision of internationally active insurance groups – update", noviembre.

Comisión de Supervisión del Sector Financiero de Luxemburgo (CSSF) (2018): "Inteligencia artificial: oportunidades, riesgos y recomendaciones para el sector financiero", diciembre.

Autoridad Monetaria de Singapur (2018): "Principios para promover la equidad, la ética, la rendición de cuentas y la transparencia (FEAT) en el uso de la inteligencia artificial y el análisis de datos en el sector financiero de Singapur", noviembre.

Asociación Nacional de Comisionados de Seguros (2020): "Principios sobre inteligencia artificial (IA)", agosto.

Netherlands Bank (DNB) (2019): "Principios generales para el uso de la Inteligencia Artificial en el sector financiero", noviembre.

Organización para la Cooperación y el Desarrollo Económicos (2019): "Principios de la OCDE sobre inteligencia artificial", mayo.

Autoridad de Regulación Prudencial (2021): "Declaración supervisora SS2/21: externalización y gestión de riesgos de terceros", marzo.

Rudin C (2019): "Deje de explicar los modelos de aprendizaje automático de caja negra para decisiones de alto riesgo y use modelos interpretables en su lugar", mayo.

The Joint Forum (2005): "Outsourcing in financial services", febrero.

Oficina del Comisionado de Información del Reino Unido (2020a): "Draft Guidance on the AI auditing framework", febrero.

——— (2020b): “Guidance on AI and data protection”, July.

Agencias reguladoras de Los Estados Unidos (2021): "Solicitud de información y comentarios sobre el uso de la IA por parte de las instituciones financieras, incluido el aprendizaje automático", marzo.

US Treasury (2018): "Un sistema financiero que crea oportunidades económicas: finanzas no bancarias, fintech e innovación", julio.

Veritas Consortium (2020): "METODOLOGÍA DE EVALUACIÓN DE PRINCIPIOS DE EQUIDAD FEAT", diciembre.

Verma S y J Rubin (2018): "Fairness definitions explained", mayo